ANDERSON'S THEOREM, THE GAUSSIAN CORRELATION INEQUALITY, AND A PROOF OF SLEPIAN'S INEQUALITY VIA THE FEYNMAN-KAC SOLUTION TO CAUCHY'S PROBLEM

A REPORT FOR STAT 542 BY CLAYTON BARNES

1. Introduction

In this report we discuss a theorem of Anderson's and trace its central use in proving inequalities for multivariate normal, and later elliptic, random variable that eventually lead naturally to the recently proven Gaussian correlation conjecture (GCC). The solution to a Cauchy problem using the Feynman-Kac formula for diffusions related to the GCC and is used to prove Slepian's inequality.

2. Anderson's Theorem

A function $f : \mathbb{R}^n \to \mathbb{R}$ is *symmetric* if $f(x) = f(-x)$ and is *unimodal* if $f^{-1}([c, \infty)) = \{x : f(x) \geq c\}$ is convex for all $c \in \mathbb{R}$. A set $K \in \mathbb{R}^n$ is symmetric if $x \in K \iff -x \in K$. In 1955, Anderson established an inequality showing the volume of a nonnegative symmetric unimodel function above a symmetric set decreases as the convex set is shifted linearly [5]. More precisely, he showed the following theorem.

Theorem 2.1 (See [5]). *Let $K \subset \mathbb{R}^n$ be symmetric and convex, and assume $f : \mathbb{R}^n \to \mathbb{R}$ is nonnegative symmetric and unimodal. Then for every fixed $y \in \mathbb{R}^n$, the function*

$$h_1(t) = \int_{K+ty} f(x)dx \tag{1}$$

is a symmetric unimodal function of $t \in \mathbb{R}$. This also implies h_1 has its maximum at $t = 0$.

Proof Sketch. First notice symmetry of h_1 follows from symmetry of f and K. To show unimodality, begin with the case $f = 1_C$ where C is symmetric and convex. Pick $0 \leq t_0 \leq t$ and $\alpha = (t_0 + t)/2t$. Notice that $C \cap (K + t_0 y) \supset \alpha[C \cap (K + ty)] + (1 - \alpha)[C \cap (K - ty)]$

because the right hand is a convex combination since $\alpha \in [.5, 1]$. Apply the Brunn-Minkowski inequality to see

$$\mu^{1/n}(C \cap (K + t_0 y)) \geq \alpha \mu^{1/n}(C \cap (K + ty)) + (1 - \alpha)\mu^{1/n}(C \cap (K - ty))$$

The sets on the right hand side have the same measure due to symmetry of both C and K. Hence

$$\mu^{1/n}(C \cap (K + t_0 y)) \geq \mu^{1/n}(C \cap (K + ty)),$$

and so $\mu(C \cap (K + t_0 y)) \geq \mu(C \cap (K + ty))$. This shows unimodality of $\int_{K+ty} 1_C(x)dx$. Now extend this to the closure of $\sum a_i 1_{C_i}$ where $a_i > 0$ and $\{C_i\}$ is a finite collection of convex symmetric sets in $\mathbb{R}^n$. It is easily seen that any symmetric unimodal f lies in this closure, proving the result. □

Fefferman, Jodeit Jr., and Perlman apply Anderson's theorem to show any linear map $A : \mathbb{R}^n \to \mathbb{R}^n$ with $\|A\| \leq 1$ will shrink a closed symmetric convex sets' intersection with the unit sphere. In other words, if $C \in \mathbb{R}^n$ is symmetric and convex, and μ denote surface measure on $S = \{x \in \mathbb{R}^n : \|x\| = 1\}$. Then

Theorem 2.2 (Theorem 1 in [1]). *$\mu(AC \cap S) \leq \mu(C \cap S)$ for any closed, convex, and symmetric set C.*

We can express $h_1(t)$ defined in (1) as

$$\int f(x)\chi_{K+ty}(x)dx = \int f(x)\chi_K(x - ty)dx = f * \chi_K(ty).$$

By Anderson's result $f * \chi_K(ty)$ is nonincreasing function of t. One can extend this to say $f_1 * f_2(ty)$ is nonincreasing as a function of t whenever f_1, f_2 are symmetric and unimodal. To prove the above theorem, the authors reduce the case to when A is replaced with $D_\lambda := \operatorname{diag}(\lambda, 1, \dots, 1)$ and $\lambda \in [0, 1]$. Let f_ϵ be an approximation to the identity, $f_\epsilon = (2\pi\epsilon)^{-n/2} \exp(-|x|^2/2\epsilon)$ and $\phi_\epsilon = \chi_C * f_\epsilon$. Consequently $\phi_\epsilon \to \chi_C$ as $\epsilon \to 0$ except perhaps at the boundary of C. Set $u_\epsilon(\lambda) = \displaystyle\int_S \phi_\epsilon(D_\lambda^{-1}x)d\mu(x)$ so that $u_\epsilon \to \displaystyle\int_S \chi_C(D_\lambda^{-1}x)d\mu(x) =$

$\mu(D_\lambda C \cap S) =: u(\lambda)$. Consider $\psi(x) = \phi_\epsilon(D_\lambda^{-1}x)$. One can show

$$u'_\epsilon(\lambda) = -\lambda^{-1}\int_S x_1 \frac{\partial \psi}{\partial x_1} d\mu(x) = -\lambda^{-1}\int_B \frac{\partial^2 \psi}{\partial x_1^2} dx$$

where x_1 is the first coordinate of x and B is the unit ball. Applying the above mentioned extension of Anderson's result shows $\chi_B * \psi$ has its maximum at the origin, and consequently $0 \geq \frac{d^2}{dt^2}\chi_B * \psi(tz)|_{t=0} = \frac{\partial^2 \chi_B * \psi}{\partial x_1^2}\Big|_{x=0}$, where $z = (1, 0, \ldots, 0)$. Combining with the above shows $u_\epsilon(\lambda) \geq 0$.

In 1970, Jogdeo used Anderson's theorem to show a two sided version of Slepian's inequality for the multivariate normal distribution [3]. Slepian's inequality states that if $X = (X_1, \ldots, X_n) \stackrel{d}{=} N(0, R)$, where $R = (\rho_{i,j}(\lambda))$, then

$$\frac{d}{d\rho_{i,j}}\mathbb{P}(X_l \leq c_l,\ l = 1, \ldots, n) \geq 0,\ c_l \in \mathbb{R}.$$

Intuitively, one expects $\mathbb{P}(|X_l| \leq c_l, l = 1, \ldots, n), c_l > 0$ to increase with the correlation of the off-diagonals. This was shown by Jogdeo in the following theorem.

Theorem 2.3 ([3])**.** *Let* $X = (X_1, \ldots, X_n) \stackrel{d}{=} N(0, R)$, *where* $R(\lambda) = \{\rho_{i,j}(\lambda)\}, \rho_{1,j}(\lambda) = \lambda\rho_{1,j},\ j > 1,\ \lambda \in [0, 1]$. *For* $c_l > 0, l = 1, \ldots, n$, *we have*

$$\frac{d}{d\lambda}\mathbb{P}(|X_l| \leq c_l,\ l = 1, \ldots, n) \geq 0.$$

Theorem 2.3 was extended to elliptically contoured distributions by Das Gupta, Eaton, Olkin, Perlman, Savage, and Sobel (or DOPESS) [2]. Let

$$\Sigma_\lambda = \begin{pmatrix} \Sigma_{11} & \lambda\Sigma_{12} \\ \lambda\Sigma_{21} & \sigma_{pp} \end{pmatrix} \in \mathbb{R}^{p \times p},$$

where Σ_{11} is $(p-1) \times (p-1)$ and $\lambda \in [0, 1]$. Let $x = (x_1, \ldots, x_p)$ be a random variable with density $|\Sigma|^{-1/2} f(x\Sigma^{-1}x^T)$ and $\int_0^\infty r^{p-1} f(r^2) dr < \infty$. This is what the authors call an elliptically contoured distribution, where they mention $y = x\Sigma^{-1/2}$ yields a density uniform on spheres and so has also been called a spherical distribution.

Theorem 2.4 (Theorem 2.1 in [2]).

$$\mathbb{P}_\lambda((x_1,\dots,x_{p-1}) \in C, |x_p| \le h),\ \ \lambda \in [0,1]$$

is nondecreasing in λ for every symmetric convex set $C \subset \mathbb{R}^{p-1}$.

Theorem 2.4 motivates the GCC in the following way: suppose we consider a general block decomposition of Σ with Σ_{12} being a $k \times p-k$ for any $1 \le k < p$, as opposed to a $(p-1)\times 1$ matrix as in Theorem 2.4. Setting $x = (x_1, x_2)$ where $x_1 \in \mathbb{R}^k$ and $x_2 \in \mathbb{R}^{p-k}$, and letting $\lambda = 0$ should minimize $\mathbb{P}_\lambda(x_1 \in C_1, x_2 \in C_2)$, where $C_1 \subset \mathbb{R}^k, C_2 \subset \mathbb{R}^{p-k}$ are symmetric convex sets. That is, one would expect

$$\mathbb{P}_\Sigma((x_1,x_2) \in (C_1,C_2)) \ge \mathbb{P}_{\overline{\Sigma}}((x_1,x_2) \in (C_1,C_2)), \tag{2}$$

as a generalization of Theorem 2.4, where $\overline{\Sigma}$ has the same main diagonals as Σ but with $\overline{\Sigma}_{12} = 0 = \overline{\Sigma}_{21}$.

2.1. **A Cauchy Problem with an Initial Condition.** Let $\sigma \in \mathbb{R}^n$ with no zero coordinates, and define $\sigma_{i,j}^2 = (\sigma\sigma^T)_{i,j}$. Denote $\Sigma = (\sigma_{i,j}^2)$. For a sufficiently smooth function $g : \mathbb{R} \to \mathbb{R}$, consider the differential operator $A_\Sigma : C^2(\mathbb{R}^n, \mathbb{R}) \to C^2(\mathbb{R}, \mathbb{R})$ which acts on g:

$$(A_\Sigma g)(x) = \frac{1}{2}\sum_{i,j} \frac{\partial^2 g}{\partial x_i \partial x_j}(x). \tag{3}$$

For a bounded f one can consider the partial differential equation solving for $u(t,x) : [0,\infty) \times \mathbb{R}^n \to \mathbb{R}$ satisfying

$$\begin{aligned} \frac{\partial}{\partial t} u(t,x) &= A_\Sigma u(t,x), \\ u(0,x) &= f(x). \end{aligned} \tag{4}$$

We will refer to (4) as the Cauchy problem with initial condition f. One can solve the Cauchy problem by averaging f composed with a certain stochastic process.

2.2. **Diffusion Solution to Cauchy Problem via Feynman-Kac Formula.** Using the same notation for $\sigma, \sigma_{i,j}^2, \Sigma, A_\Sigma$ in the Cauchy problem above, let $X(t)$ be an n dimensional diffusion defined by

$$dX(t) = \sigma dB(t),$$

where $\sigma \in \mathbb{R}^n$ and $B(t)$ is a standard n dimensional Brownian motion. At a fixed time t_0, $X(t_0) \stackrel{d}{=} N(0, t_0\Sigma)$.

Theorem 2.5 (See chapter 4.4 in [4]). *Define*

$$u(t, x) = \mathbb{E}(f(X(t))|X(0) = x),$$

then $u(t, x)$ solves the Cauchy problem (4).

Here $\mathbb{E}(f(X(t))|X(0) = x)$ is the expected value of $f(X(t))$ conditioned on $X(0) = x$. That is, X begins at x. The Feynman-Kac formula is typically written by choosing a $T > 0$ and setting $z(t, x) = \mathbb{E}(f(X(T))|X(t) = x), t \in [0, T]$, which solves

$$\begin{aligned} \frac{\partial}{\partial t} z(t, x) &= -A_\Sigma z(t, x), \\ z(T, x) &= f(x). \end{aligned} \tag{5}$$

Somtimes this is called the Kolmogorov backward equation. In any case, a change of variable $u(t, x) := z(T - t, x)$ demonstrates u solves (4) for $t \in [0, T]$.

Let $f(x_1, \ldots, x_n) = \prod_{i=1}^n f_i(x_i)$ where $f_i : \mathbb{R} \to \mathbb{R}$ are defined to be smooth bump function. For example, one may take

$$f_i(t) = g(t/c_i), \;\; c_i \neq 0$$

where $g : \mathbb{R} \to \mathbb{R}$ is a smooth symmetric bump function, or $f_i(t) = g(t + c_i)$. (We take a particular form of g later for the corollary demonstrating Slepian's inequality). In other words, $f(x_1, \ldots, x_n)$ is approximately the indicator of an n dimensional centered rectangle. Consider $(\sigma_{i,j}^2) = \Sigma$ and $(\overline{\sigma}_{i,j}^2) = \overline{\Sigma}$, where $\overline{\sigma}_{i,j}^2 = 0$ whenever $1 \leq i \leq k < j \leq n$ (and extend this symmetrically for $\overline{\sigma}_{j,i}^2$) for some given $1 \leq k < n$. Otherwise let $\overline{\sigma}_{i,j}^2$ agree with $\sigma_{i,j}^2$. We further assume $\sigma_{i,j}^2 \geq 0$ when $1 \leq i \leq k < j \leq n$.

In other words, we are decorrelating the coordinates of $X(t)$ up to k from the coordinates after k. Now $A := A_\Sigma, A_{\overline{\Sigma}} =: B$ both define a Cauchy problem with initial condition f.

Denote $u(t,x), v(t,x)$ as the Cauchy problem associated with A, B respectively. By taking a sequence of bump functions converging to the indicator, equation (2) is morally equivalent to

$$u(1,0) \geq v(1,0)$$

when C_1, C_2 are k and $n-k$ dimensional centered rectangles.

Recall the definition of the differential operators A and B in (3), and note that

$$C = A - B = \sum_{1 \leq i \leq k < j \leq n} \sigma_{i,j}^2 \frac{\partial^2}{\partial x_i \partial x_j},$$

where the 1/2 drops out because each partial shows up twice due to equivalence of mixed partials. Define $w(t,x) = u(t,x) - v(t,x)$. Because u, v are solutions to their respective Cauchy problems, linearity of the operators A, B, C give

$$\begin{aligned}
\frac{\partial}{\partial t} w(t,x) &= \frac{\partial}{\partial t}(u(t,x) - v(t,x)) \\
&= Au(t,x) - Bv(t,x) \\
&= (B+C)u(t,x) - Bv(t,x) \\
&= B(u(t,x) - v(t,x)) + Cu(t,x) \\
&= Bw(t,x) + Cu(t,x),
\end{aligned}$$

while the initial condition is

$$w(0,x) = u(0,x) - v(0,x) = 0.$$

If we can show $Cu(t,x) \geq 0$, then it follows immediately that

$$\frac{\partial}{\partial t} w(t,x) \geq Bw(t,x), \;\; w(0,x) = 0.$$

Consequently, $w(t,x)$ would be dominated below from the solution to the Cauchy problem with operator B and an initial condition of the zero function, yielding $0 \leq w(1,0) = u(1,0) - v(1,0)$ in particular.

Proposition 2.6. *Let $p_t(x,y)$ be the transition density of X with $\Sigma = (\sigma_{i,j}^2)$ given as above, and recall $f(z) = \prod_{i=1}^n f_i(z_i)$. Then*

$$C_x u(t,x) = \int (C_y f)(y) p_t(x,y)dy = \sum_{1\le i\le k<j} \sigma_{i,j}^2 \int \left(\prod_{i\neq l\neq j} f_i(y_i)\right) f_i'(y_i)f_j'(y_j)p_t(x,y)dy \tag{6}$$

For clarity, C_y means the operator C acts on the variable $y = (y_1,\dots,y_n)$.

Sketch. By definition of the transition density $p_t(x,y)$, $u(t,x) = \mathbb{E}(f(X(t))|X(0) = x) = \int f(y)p_t(x,y)dy$. Consequently

$$C_x u(t,x) = C_x \int f(y)p_t(x,y)dy = \int f(y)C_x p_t(x,y)dy.$$

However, transitioning from x to y in a time of t is the same as transitioning from the origin to $y-x$ in a time of t. So $p_t(x,y) = p_t(0,y-x)$. Because the operator C is a sum of mixed partials $\partial^2/\partial x_i\partial x_j$ with $i\neq j$, and because $p_t(0,y-x) = |t\Sigma|^{-n/2}\exp(-\frac{1}{2t}(y-x)\Sigma^{-1}(y-x)^T)$, the quadratic nature of the exponent allows one to show $\partial^2 p_t(0,y-x)/\partial x_i\partial x_j = \partial^2 p_t(0,y-x)/\partial y_i\partial y_j$. Therefore,

$$C_x u(t,x) = \int f(y)C_y p_t(x,y)dy.$$

Now apply an integration by parts for each mixed partial in C_y to move this operator from $p_t(0,y-x)$ to $f(y)$, giving

$$C_x u(t,x) = \int (C_y f)(y)p_t(x,y)dy.$$

This reduces to the right hand side of (6) in the case $f(z) = \prod_{i=1}^n f(z_i)$. □

The above proposition yields Slepian's Inequality as a corollary

Corollary 2.7 (Slepian's Inequality)**.** *In the notation $\Sigma = (\sigma_{i,j}^2)$ and $\overline{\Sigma} = (\overline{\sigma}_{i,j}^2)$ defined above, $\mathbb{P}_\Sigma(X_1\le c_1,\dots,X_n\le c_n) \ge \mathbb{P}_{\overline{\Sigma}}(X_1\le l_1,\dots,X_n\le l_n)$.*

Proof. Let $f_i(z) = g(z-l_i)$ be smooth, increasing (thinking of this function as approximating $1_{(-\infty,l_i]}$). This follows by taking the g to approximate $1_{(-\infty,0]}$. Then f_i', f_i are both nonnegative, and consequently the right hand side of (6) is nonnegative when adding our assumption that $\sigma_{i,j}^2 \geq 0$ for $1 \leq i \leq k < j \leq n$. Using argument preceeding the proposition, $w(t,x) \geq 0$. $\square$

Remark 2.8 (Some Rambling Thoughts)**.** I have the feeling that I'm not constructing a new argument. And it's possible the proposition is not quite correct. But if it is, this is where the Brascamp-Lieb inequality, or some theorem showing the right hand side of (6) is nonnegative for all x (or just $x = 0$?), would come into play. This right hand side can be written as a convolution of $\left(\prod_{i \neq l \neq j} f_i(y_i)\right) f_i'(y_i) f_j'(y_j)$ with $p_t(0,y)$ since $p_t(0,y-x) = p_t(x,y)$.

It seems to me that we only need the RHS of (6) to be nonnegative at $x = 0$. (Perhaps this is not the case.)

However, the reason I doubt the validity of the above proposition is because I obtain zero for the right hand side when f_i is a smooth and symmetric bump function "approximating" the indicator $1_{[-c_i,c_i]}$ when $x = 0$. This occurs because $f_i'(y_i) f_j'(y_j)$ is symmetric but negative in the orthant with e_i and $-e_j$ "pointing inside" or $-e_i$ and e_j pointing inside, (with e_1, e_2 this would be the second and fourth quadrant). Consequently, setting $x = 0$ would say the right hand side of (6) is zero (at $x = 0$). This would say that $w(t,0) = 0$ if we look carefully at the inequalities: One would have $w(t,0)$ solve $\partial_t w(t,0) = Bw(t,0)$ with $w(0,0) = 0$. In other words, $u(t,0) = v(t,0)$ which must not always be true? I think there is something fishy with either (1) the proposition above, (2) the argument with needing $Bw(t,x) + Cu(t,x) \geq Bw(t,x)$ implying $w(t,x) \geq 0$ or (3) I'm mistaken about the RHS of (6) being zero when f_i, f_j are symmetric bump functions estimating the $1_{[-c_i,c_i]}$. Or we really do need the inequality for all x.

If the above proposition is false, then we could simply let C act on $p_t(0,y-x)$, which is not too bad either. But we would eventually we need to this is nonnegative.

References

[1] Charles Fefferman, Max Jodeit, and Michael D Perlman. A spherical surface measure inequality for convex sets. *Proceedings of the American Mathematical Society*, 33(1):114–119, 1972.

[2] S Das Gupta, Morris L Eaton, Ingram Olkin, M Perlman, Leonard J Savage, and Milton Sobel. Inequalities on the probability content of convex regions for elliptically contoured distributions. In *Proceedings*

of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), volume 2, pages 241–265, 1972.

[3] Kumar Jogdeo. A simple proof of an inequality for multivariate normal probabilities of rectangles. *The Annals of Mathematical Statistics*, 41(4):1357–1359, 1970.

[4] I. Karatzas and S. Shreve. *Brownian Motion and Stochastic Calculus*. Springer-Verlag, 2nd edition, 1991.

[5] Michael D. Perlman. Anderson's Theorem on the Integral of a Symmetric Unimodal Function over a Symmetric Convex Set, and Its Applications in Probability and Statistics. 2007.